Human echolocation: 2D shape discrimination using features extracted from acoustic echoes

Xuelian Yu✉, Lore Thaler, Christopher J. Baker, Graeme E. Smith and Linsen Zhao

Some blind humans have developed the ability to perceive their silent surrounding by using echolocation based on tongue clicks. Past research has also shown that blind echolocators can use information gained from multiple echoic 'views', provided through head movements, to successfully identify 2D shapes. Here, echo features that might be used by blind humans to discriminate 2D shapes are investigated. Echoes from four shapes are collected with a custom-built acoustic radar and various features are extracted. By piecing together individual features across the measurement plane, it is found that total power and spectral centroid are two salient features for shape discrimination from multiple echoic views.

Introduction: Like echolocating bats and dolphins, some blind humans are capable of echolocation. By making a click with their tongue and listening to returning echoes, they can gather information about their environment such as range, size, shape, and texture of objects [1, 2].

Although human echolocation is receiving increasing research attention, comparably little research has been dedicated to understand the ability of blind people to discriminate 2D shapes via echolocation. Rice [3], in 1967, showed that blind echolocators could distinguish between a circle, square, and triangle. In 2014, Milne *et al.* [4] showed that blind echolocators can successfully determine the shape of objects with different contours only, and that head movements were essential. Specifically, when echolocators kept their head still, their ability to correctly determine shape was at chance levels. When they were allowed to move their heads, they performed with 80% accuracy. Thus, it is reasonable to assume that blind echolocators determine the shape of objects by integrating acoustic information across multiple echoic views. What remains unclear is which acoustic echo features might be used during this process.

This Letter provides insights useful for exploring the above. We collect echoes by spatially sampling with a custom-built acoustic radar and then calculate features. The emissions we used were human inspired synthetic clicks developed in [5]. Among various features, the total power (TP) and spectral centroid (SC) provided the most meaningful shape discrimination performance. These results contribute to our understanding of human echolocation for 2D shape discrimination and can be used to guide future research.

Tongue click transmissions: We use the tongue clicks developed in [5] based on measurements from three blind echolocators (EE1, EE2, and EE3). Fig. 1, top row, illustrates waveforms of the three clicks used as transmissions, each from a single echolocator. The middle and bottom rows show the power spectral density (PSD) and spectrogram of each click, respectively. It can be seen that the main frequency components for EE1–EE3 are all within 2–4 kHz range and all clicks have high-frequency components at around 10-11 kHz.

Data collection: To replicate the conditions used with human echolocators [4], we used four 2D shapes: (i) 40×40 cm^2 square; (ii) 52 cm equilateral triangle; (iii) 100×16 cm^2 rectangle; and (iv) 16×100 cm^2 rectangle. All the shapes were flat and only differed in their contour. They were made of smooth cardboard, 0.5 cm thick.

Experimental data was collected with a custom-built setup and the main part of it was an acoustic radar. A loudspeaker (KRK Systems ROKIT 6 Powered Monitor) and two condenser microphones (AKG Perception 170) form the radar head, imitating the human head 'mouth-ears' arrangement. With three perpendicular sliders, the radar head is able to make left-and-right, up-and-down, and back-and-forth motions, coarsely imitating the human head movements. The loudspeaker and microphones are connected with an audio interface (RME Fireface UC, 96 kHz, 24 bit) for click transmission and echo reception. A desk computer with custom-written software was used to control slider motion and data collection.

Data collection took place in an echo-dampened room (3.5 m deep, 3.1 m wide, 3.1 m high, four walls, and ceiling covered in foam sheets, the floor covered in damping material). On each measurement, one of the four shapes was presented. The shape was positioned fixed with its surface perpendicular to the floor, and facing the radar setup at 80 cm or 40 cm distance. Fig. 2 (left) shows a schematic drawing of the experimental setup.

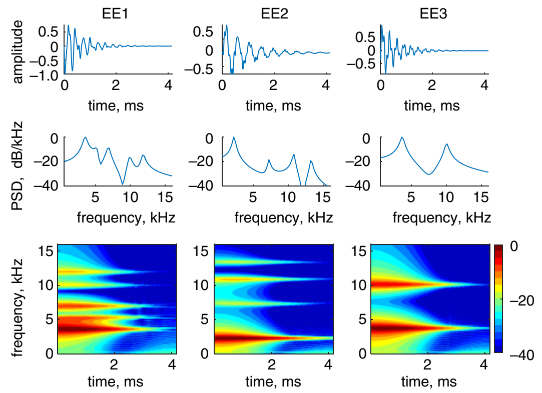

Fig. 1 *Tongue click transmissions*

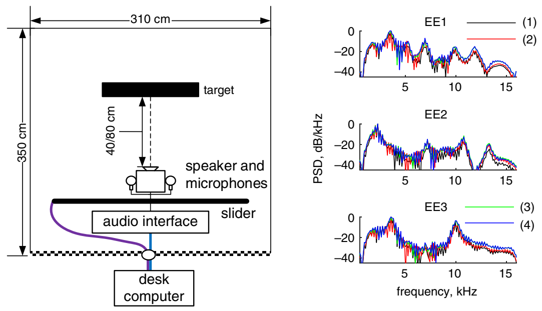

Fig. 2 *Left: experimental setup. Right: PSDs of echoes from centre of each shape at 80 cm: (i) square, (ii) tri., (iii) horiz. rect., and (iv) vert. rect.*

At 80 cm distance, one measurement was taken for each shape with the radar pointing at the centre. This corresponds to the case of 'fixed-position' conditions in [4] when participants were asked to remain still, i.e. to not move their head. Note that [4] had also done a 'fixed-position' control condition at 40 cm distance, but we here used 80 cm because echolocators reported that they found this distance more useful (even though they performed at the chance at both distances).

At 40 cm distance, the radar started from the centre of shape and moved with $2\text{cm} \times 2$ cm steps in the horizontal–vertical plane (parallel to the surface of shape). One measurement was taken at each step so as to form a total of $M \times N$ echo views of each shape, where M and N are the total steps that the radar moved horizontally and vertically, respectively. The shape was placed such that the centre of shape was at the centre of the measurement plane. This group of data collection was deliberately designed to correspond to the case of 'free-moving' conditions in [4] when participants were allowed to make head movements.

Data analysis and results corresponding to fixed-position conditions: PSDs were calculated for each click and shape. Although we recorded data with two microphones, only one channel was used for analysis because the results from the two channels were the same.

Fig. 2 (right) presents PSDs of echoes collected from the centre of each shape at 80 cm distance. It can be seen that spectral characteristics of echoes are largely determined by the spectral characteristic of the clicks (i.e. compare Fig. 1), and that for each click PSDs are very similar across the four shapes. Therefore, we can speculate that it is very difficult to distinguish the four shapes by using echoes collected from only the centre of each shape. This is consistent with the finding in [4], where echolocators were unable to discriminate object shapes in fixed-position conditions.

Data analysis and results corresponding to free-moving conditions: As reported in [4], blind expert echolocators correctly and reliably indicated 2D shapes of objects in free-moving conditions. It was suggested that making head movements while producing tongue clicks would allow a person to accumulate echo-acoustic information across multiple views, which are then pieced together by the brain to create an overall perceptual representation of 2D contour shape.

Aiming to determine which features might convey useful shape identification information or could possibly provide some perceptual representation of object shape, a variety of features were extracted from the echoes at each sample point to obtain 2D feature plots. The following four features were considered:

(i) TP in decibels (dB), calculated by summing the values of the PSD of echo from 500 Hz to 16 kHz.
(ii) SC in hertz, the centroid of the mass of the PSD of echo from 500 Hz to 16 kHz.
(iii) Main frequency power (MFP) in dB, defined as the power within 2–4 kHz. It is supposed that these strongest frequency contents might indicate the object shape.
(iv) High-frequency power (HFP) in dB, defined as the power within 10–11 kHz range. It is believed that high frequencies help to reveal fine information which might be useful for shape identification.

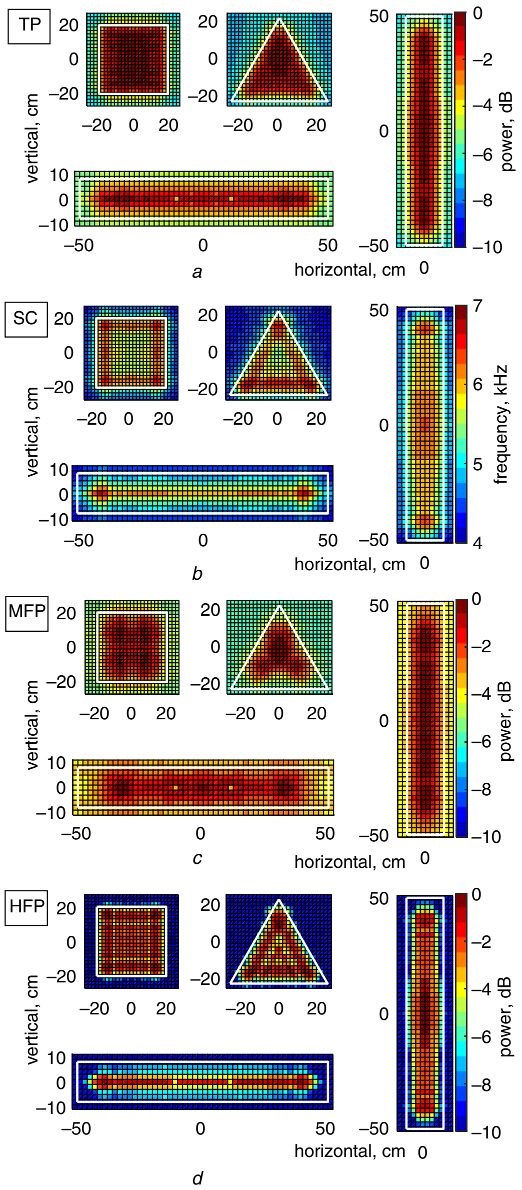

Fig. 3 *Feature plots averaged across three clicks. White lines: shape contours*

a TP plots
b SC plots
c MFP plots
d HFP plots

Fig. 3 plots the four features of each shape, each plot being an averaged result across three clicks. For illustration, we also plot the physical contour of each shape in Fig. 3. As can be seen, all feature plots could provide a meaningful perceptual representation of objects. Intuitively, the TP and SC plots show that changes in TP and SC correctly follow the shape contour suggesting that both may be employed for consistent and reliable shape discrimination. By comparison, the MFP and HFP plots show less defined contours, though, they are still enough to distinguish among different shapes.

The results indicate that TP and SC might be two salient features for 2D contour shape discrimination. This conclusion is consistent with a study of acoustic cues used by humans listening to dolphin sounds to discriminate among object shapes [6]. DeLong *et al.* [6] proposed that the overall loudness and overall pitch were the most important cues for making a decision between shapes. These properties are comparable with TP and SC, respectively.

Comparing further the TP and SC plots of the square in Figs. 3*a* and *b*, we can find that TP obtains its maximum, almost uniformly, in areas inside the shape edges (taking into consideration unavoidable measurement errors) and drops outside, whereas SC obtains its maximum in areas along the shape contour and drops inside. Anecdotal evidence based on questioning some blind expert echolocators suggests that during the shape identification task, they firstly tried to find the edges of the object and then work their way in between to trace the contour of the object. On the basis of our results, it would seem possible that this was achieved by more easily using TP and SC. If this was the case, we would further expect those areas that carry the most information (i.e. differences in TP or SC as one moves from one sample to the next) might be areas in space that blind humans explore with their head movements during echolocation. We will address these questions in future research.

Conclusion: In this Letter, we developed an experimental setup and an analysis method to illustrate echo features that may be used in human echolocation for 2D shape discrimination. We found that TP and SC are two salient features for shape discrimination, a conclusion consistent with [6]. Plots of features across the measurement plane provide useful perceptual representations of objects. Furthermore, a combination of both features may enable successful shape discrimination. The results of this work can be used to guide future research on cognitive processes used by echolocating humans that may be used for future cognitive radar systems.

Submitted: *26 February 2018*
doi: 10.1049/el.2018.0680
One or more of the Figures in this Letter are available in colour online.

Xuelian Yu and Linsen Zhao (*School of Information and Communication, University of Electronic Science and Technology of China, Chengdu 61173, People's Republic of China*)

✉ E-mail: xlyu2012@uestc.edu.cn

Lore Thaler (*Department of Psychology, Durham University, Durham, United Kingdom*)

Christopher J. Baker (*Department of Electronic Electrical and Systems Engineering, University of Birmingham, Birmingham, United Kingdom*)

Graeme E. Smith (*Department of Electrical and Computer Engineering, Ohio State University, Columbus, OH 43210, USA*)

References

1 Kolarik, A.J., Cirstea, S., Pardhan, S., *et al.*: 'A summary of research investigating echolocation abilities of blind and sighted humans', *Hear. Res.*, 2014, **310C**, pp. 60–68
2 Thaler, L., and Goodale, M.A.: 'Echolocation in humans: an overview', *WIREs Cogn. Sci.*, 2016, **7**, pp. 382–393, doi: 10.1002/wcs.1408
3 Rice, C.E.: 'Human echo perception', *Sci.*, 1967, **155**, pp. 656–664
4 Milne, J.L., Goodale, M.A., and Thaler, L.: 'The role of head movements in the discrimination of 2-D shape by blind echolocation experts', *Atten. Percept. Psychophys*, 2014, **76**, (6), pp. 1828–1837
5 Thaler, L., Reich, G.M., Zhang, X.Y., *et al.*: 'Mouth-clicks used by blind human echolocators – signal description and model based signal synthesis', *PLOS Comput. Biol.*, 2017, **13**, (8), p. e1005670, doi: 10.1371/journal.pcbi.1005670
6 DeLong, C.M., Au, W.W.L., Harley, H.E., *et al.*: 'Human listeners provide insights into echo features used by dolphins (*Tursiops truncatus*) to discriminate among objects', *J. Comp. Psychol.*, 2007, **121**, (3), pp. 306–319, doi:10.1037/0735-7036.121.3.306